

25. Тузик А.И. Дискретные уравнения типа свертки с почти стабилизирующимися множителями специального вида // Тезисы докл. междунар. конф. "Аналитические методы

анализа и дифференц. уравн." – Мн.: БГУ, 2001. С. 162 – 163.

УДК 681.324

Гладкий И.И., Головкин В.А., Махнист Л.П.

ОБУЧЕНИЕ НЕЙРОННЫХ СЕТЕЙ С ИСПОЛЬЗОВАНИЕМ МЕТОДА НАИСКОРЕЙШЕГО СПУСКА

1. ВВЕДЕНИЕ

Рассмотрим нейронную сеть, состоящую из n нейронных элементов распределительного слоя и m - выходного слоя (рисунок 1).

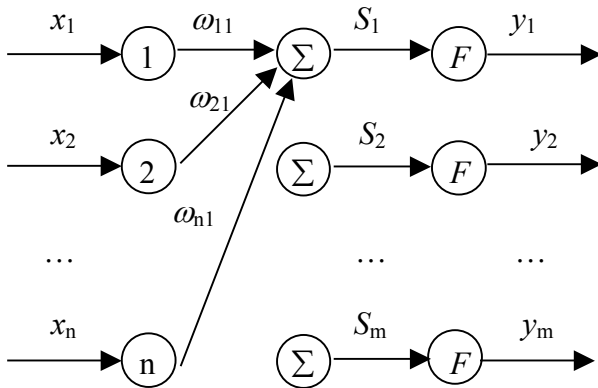


Рисунок 1 – Схема функционирования нейронной сети.

Для данной сети каждый нейрон распределительного слоя имеет синаптические связи со всеми нейронами обрабатывающего слоя. В качестве нейронов выходного слоя используются элементы с некоторой функцией активации F .

Рассмотрим наиболее распространенные дифференцируемые на всей числовой прямой функции активации, их производные первого и второго порядка:

- 1). Линейная функция: $F(S) = cS$, $F(0) = 0$,
 $F'(S) = c$, $F'(0) = c$,
 $F''(S) = 0$, $F''(0) = 0$.

- 2). Сигмоидная функция с областью значений

$$E(F) = (0; 1): F(S) = \frac{1}{1 + e^{-cS}}, F(0) = \frac{1}{2},$$

$$F'(S) = \frac{ce^{-cS}}{(1 + e^{-cS})^2} = \frac{c}{1 + e^{-cS}} \cdot \left(1 - \frac{1}{1 + e^{-cS}}\right) = cF(S)(1 - F(S)),$$

$$F'(0) = \frac{c}{4},$$

$$F''(S) = cF'(S) - 2cF(S)F'(S) = cF'(S)(1 - 2F(S)) = c^2F(S)(1 - F(S))(1 - 2F(S)),$$

$$F''(0) = 0.$$

- 3). Биполярная сигмоидная функция с областью значений $E(F) = (-1; 1)$:

$$F(S) = \frac{2}{1 + e^{-cS}} - 1, F(0) = 0,$$

$$F'(S) = \frac{2c \cdot e^{-cS}}{(1 + e^{-cS})^2} = \frac{c}{2} \cdot \left(1 + \left(\frac{2}{1 + e^{-cS}} - 1\right)\right) \times$$

$$\times \left(1 - \left(\frac{2}{1 + e^{-cS}} - 1\right)\right) = \frac{c}{2}(1 - F^2(S)),$$

$$F'(0) = \frac{c}{2},$$

$$F''(S) = -cF(S)F'(S) = -\frac{c^2}{2}F(S)(1 - F^2(S)),$$

$$F''(0) = 0.$$

- 4). Функция распределения Коши с областью значений $E(F) = (0; 1)$:

$$F(S) = \frac{1}{2} + \frac{1}{\pi} \arctg(cS), F(0) = \frac{1}{2},$$

$$F'(S) = \frac{c}{\pi(1 + (cS)^2)} = \frac{c}{\pi \left(1 + \tg^2\left(\pi F(S) - \frac{\pi}{2}\right)\right)},$$

$$F'(0) = \frac{c}{\pi}.$$

$$F''(S) = -\frac{2c^3S}{\pi(1 + (cS)^2)^2} = -\frac{2c^2 \tg\left(\pi F(S) - \frac{\pi}{2}\right)}{\pi \left(1 + \tg^2\left(\pi F(S) - \frac{\pi}{2}\right)\right)^2},$$

$$F''(0) = 0.$$

- 5). Обратная тригонометрическая функция арктангенс с областью значений $E(F) = (-1; 1)$:

Гладкий Иван Иванович. Ассистент кафедры высшей математики Брестского государственного технического университета.

Головкин Владимир Адамович. Докторант ИТК НАНБ, к.т.н., доцент, профессор каф. ЭВМиС Брестского государственного технического университета.

Махнист Леонид Петрович. К.т.н., доцент каф. высшей математики Брестского государственного технического университета.

Беларусь, БГТУ, 22017, г. Брест, ул. Московская, 267.

Физика, математика, химия

$$F(S) = \frac{2}{\pi} \operatorname{arctg}(cS), \quad F(0) = 0,$$

$$F'(S) = \frac{2c}{\pi(1+(cS)^2)} = \frac{2c}{\pi\left(1+\operatorname{tg}^2\left(\frac{\pi}{2}F(S)\right)\right)},$$

$$F'(0) = \frac{2c}{\pi}.$$

$$F''(S) = -\frac{4c^3S}{\pi(1+(cS)^2)^2} = -\frac{4c^2\operatorname{tg}\left(\frac{\pi}{2}F(S)\right)}{\pi\left(1+\operatorname{tg}^2\left(\frac{\pi}{2}F(S)\right)\right)^2},$$

$$F''(0) = 0.$$

6). Гиперболический тангенс с областью значений $E(F) = (0; 1)$:

$$F(S) = \frac{1}{2}(th(cS) + 1) = \frac{e^{cS}}{e^{cS} + e^{-cS}}, \quad F(0) = \frac{1}{2},$$

$$F'(S) = \frac{c}{2ch^2(cS)} = \frac{c}{2} \cdot \frac{ch^2(cS) - sh^2(cS)}{ch^2(cS)} =$$

$$= \frac{c}{2}(1 - th^2(cS)) = \frac{c}{2}(1 + th(cS))(2 - th(cS) - 1) =$$

$$= \frac{c}{2} \cdot 2F(S)(2 - 2F(S)) = 2cF(S)(1 - F(S)), \quad F'(0) = \frac{c}{2}.$$

$$F''(S) = 2cF'(S)(1 - 2F(S)) =$$

$$= 4c^2F(S)(1 - F(S))(1 - 2F(S)),$$

$$F''(0) = 0.$$

7). Гиперболический тангенс с областью значений $E(F) = (-1; 1)$:

$$F(S) = th(cS) = \frac{e^{cS} - e^{-cS}}{e^{cS} + e^{-cS}}, \quad F(0) = 0,$$

$$F'(S) = \frac{c}{ch^2(cS)} = c \cdot \frac{ch^2(cS) - sh^2(cS)}{ch^2(cS)} =$$

$$= c(1 - th^2(cS)) = c(1 - F^2(S)),$$

$$F'(0) = c,$$

$$F''(S) = -2cF(S)F'(S) = -2c^2F(S)(1 - F^2(S)),$$

$$F''(0) = 0.$$

8). Обратная гиперболическая функция аресинус:

$$F(S) = \operatorname{arsh}(cS) = \ln\left(cS + \sqrt{(cS)^2 + 1}\right), \quad F(0) = 0,$$

$$F'(S) = \frac{c + \frac{c^2S}{\sqrt{(cS)^2 + 1}}}{cS + \sqrt{(cS)^2 + 1}} = \frac{c}{\sqrt{(cS)^2 + 1}} = \frac{c}{e^{F(S)} - cS},$$

$$F'(0) = c,$$

$$F''(S) = -\frac{c}{(e^{F(S)} - cS)^2} \cdot (e^{F(S)}F'(S) - c) =$$

$$= -\frac{c}{(e^{F(S)} - cS)^2} \cdot \left(e^{F(S)} \frac{c}{e^{F(S)} - cS} - c\right) = -\frac{c^3S}{(e^{F(S)} - cS)^3}$$
или $F''(S) = -\frac{c^3S}{\sqrt{((cS)^2 + 1)^3}}, \quad F''(0) = 0.$

Легко доказать следующее утверждение, связывающее, например, производные n -ого порядка функций активации, приведенные в пунктах 2 и 3, 4 и 5, 6 и 7.

Утверждение. Пусть $F(S)$ – дифференцируемая функция активации с областью значений на интервале $(a; b)$. Тогда дифференцируемая функция

$$G(S) = \frac{\beta - \alpha}{b - a}(F(S) - a) + \alpha$$

является функцией активации с областью значений на интервале $(\alpha; \beta)$ и имеют место соотношения для производных функций $F(S)$ и $G(S)$ n -го

$$\text{порядка: } G^{(n)}(S) = \frac{\beta - \alpha}{b - a} F^{(n)}(S).$$

Замечание. Если известна функция H такая, что $F^{(n)}(S) = H(F(S))$, то

$$G^{(n)}(S) = \frac{\beta - \alpha}{b - a} \cdot H\left(\frac{b - a}{\beta - \alpha}(G(S) - \alpha) + a\right).$$

Данное утверждение позволяет использовать, вместо процедуры масштабирования выходных данных нейронной сети [1], построение функции активации с областью значений, которой принадлежат выходные данные нейронной сети.

2. ВЫБОР АДАПТИВНОГО ШАГА ОБУЧЕНИЯ

Получим выражение для адаптивного шага обучения нейронной сети после подачи на вход сети нескольких образов $\mathbf{x}^k = (x_1^k, \dots, x_n^k)$ ($k = \overline{1, L}$).

Теорема. Для нейронной сети величина адаптивного шага обучения $\alpha(t)$ в момент времени t определяется соотношением:

$$\alpha(t) = \frac{\sum_{k=1}^L \sum_{j=1}^m (y_j^k(t) - t_j^k) F'(S_j^k(t)) a_j^k(t)}{\sum_{k=1}^L \sum_{j=1}^m \left((F'(S_j^k(t)))^2 + (y_j^k(t) - t_j^k) F''(S_j^k(t)) \right) (a_j^k(t))^2} \quad (1)$$

$$\text{где } a_j^k(t) = \sum_{p=1}^L (y_j^p(t) - t_j^p) F'(S_j^p(t)) \left(\sum_{i=1}^n x_i^p x_i^k + 1 \right),$$

$$j = \overline{1, m}, \quad k = \overline{1, L}.$$

Модификация синаптических связей с использованием адаптивного шага обучения определяется выражениями:

$$\omega_{ij}(t+1) = \omega_{ij}(t) - \alpha(t) \cdot \sum_{k=1}^L (y_j^k(t) - t_j^k) \cdot F'(S_j^k(t)) x_i^k,$$

$$i = \overline{1, n}, \quad j = \overline{1, m} \quad (2)$$

$$T_j(t+1) = T_j(t) + \alpha(t) \cdot \sum_{k=1}^L (y_j^k(t) - t_j^k) F'(S_j^k(t)),$$

$$j = \overline{1, m}. \quad (3)$$

Доказательство: Выходное значение j -ого нейрона сети для k -ого образа в момент времени t определяется выражением:

$$y_j^k(t) = F(S_j^k(t)),$$

где

$$S_j^k(t) = \sum_i \omega_{ij}(t) x_i^k - T_j(t), \quad j = \overline{1, m}, \quad k = \overline{1, L}. \quad (4)$$

Среднеквадратичная ошибка сети для всей обучающей выборки при групповом обучении равна:

$$E_S(t) = \sum_k \sum_j E_j^k(t) = \frac{1}{2} \sum_k \sum_j (y_j^k(t) - t_j^k)^2,$$

где t_j^k - эталонное выходное значение j -ого нейрона сети для k -ого образа.

Для определения адаптивного шага обучения с использованием метода наискорейшего спуска необходимо определить такой шаг обучения сети, который минимизирует среднеквадратичную ошибку $E_S(t+1) = \sum_k \sum_j E_j^k(t+1)$.

Подставляя выражения для изменения весовых коэффициентов

$$\omega_{ij}(t+1) = \omega_{ij}(t) - \alpha(t) \frac{\partial E_S}{\partial \omega_{ij}(t)}, \quad i = \overline{1, n}, \quad j = \overline{1, m},$$

и порогов нейронных элементов

$$T_j(t+1) = T_j(t) - \alpha(t) \frac{\partial E_S}{\partial T_j(t)}, \quad j = \overline{1, m},$$

в соотношение (4) для момента времени $t+1$, получим:

$$S_j^k(t+1) = \sum_i \left(\omega_{ij}(t) - \alpha(t) \frac{\partial E_S}{\partial \omega_{ij}(t)} \right) x_i^k - \left(T_j(t) - \alpha(t) \frac{\partial E_S}{\partial T_j(t)} \right),$$

или

$$S_j^k(t+1) = \sum_i \omega_{ij}(t) x_i^k - T_j(t) - \alpha(t) \cdot \left(\sum_i \frac{\partial E_S}{\partial \omega_{ij}(t)} x_i^k - \frac{\partial E_S}{\partial T_j(t)} \right),$$

$$j = \overline{1, m}, \quad k = \overline{1, L}$$

Обозначим:

$$a_j^k(t) = \sum_i \frac{\partial E_S}{\partial \omega_{ij}(t)} x_i^k - \frac{\partial E_S}{\partial T_j(t)}, \quad j = \overline{1, m}, \quad k = \overline{1, L}. \quad (5)$$

Тогда

$$S_j^k(t+1) = S_j^k(t) - \alpha(t) \cdot a_j^k(t), \quad j = \overline{1, m}, \quad k = \overline{1, L}. \quad (6)$$

Ограничиваясь тремя членами разложения в ряд Тейлора в окрестности некоторой точки $S_j^k(t)$ для ошибки сети $E_j^k(t+1)$ j -ого нейрона сети для k -ого образа, имеем

$$E_j^k(t+1) = \frac{1}{2} (y_j^k(t+1) - t_j^k)^2 =$$

$$= \frac{1}{2} (y_j^k(t) - t_j^k)^2 + (y_j^k(t) - t_j^k) F'(S_j^k(t)) (S_j^k(t+1) - S_j^k(t)) +$$

$$+ \frac{1}{2} \left((F'(S_j^k(t)))^2 + (y_j^k(t) - t_j^k) F''(S_j^k(t)) \right) (S_j^k(t+1) - S_j^k(t))^2 \quad (7)$$

Подставляя (6) в последнее соотношение, получим

$$E_j^k(t+1) = \frac{1}{2} (y_j^k(t) - t_j^k)^2 + 2(y_j^k(t) - t_j^k) F'(S_j^k(t)) (-\alpha(t) \cdot a_j^k(t)) +$$

$$+ \left((F'(S_j^k(t)))^2 + (y_j^k(t) - t_j^k) F''(S_j^k(t)) \right) (-\alpha(t) \cdot a_j^k(t))^2$$

или в сокращенном виде

$$E_j^k(t+1) = \frac{1}{2} (y_j^k - t_j^k)^2 - (y_j^k - t_j^k) F'(S_j^k) a_j^k \times$$

$$\times \alpha + \frac{1}{2} \left((F'(S_j^k))^2 + (y_j^k - t_j^k) F''(S_j^k) \right) (a_j^k)^2 \cdot \alpha^2$$

Для определения адаптивного шага обучения необходимо найти такое значение α , чтобы среднеквадратичная ошибка была минимальной:

$$\min_{\alpha} E_S(t+1) = \sum_k \sum_j E_j^k(t+1) =$$

$$= \min_{\alpha} \frac{1}{2} \sum_k \sum_j (y_j^k(t+1) - t_j^k)^2 =$$

$$= \min_{\alpha} \sum_k \sum_j \left(\frac{1}{2} (y_j^k - t_j^k)^2 - (y_j^k - t_j^k) F'(S_j^k) a_j^k \cdot \alpha + \right.$$

$$\left. + \frac{1}{2} \left((F'(S_j^k))^2 + (y_j^k - t_j^k) F''(S_j^k) \right) (a_j^k)^2 \cdot \alpha^2 \right)$$

Тогда

$$\frac{\partial E_S}{\partial \alpha} = \sum_k \sum_j \left(- (y_j^k - t_j^k) F'(S_j^k) a_j^k + \right.$$

$$\left. + \left((F'(S_j^k))^2 + (y_j^k - t_j^k) F''(S_j^k) \right) (a_j^k)^2 \cdot \alpha \right) = 0$$

если

$$\alpha = \frac{\sum_k \sum_j (y_j^k - t_j^k) F'(S_j^k) a_j^k}{\sum_k \sum_j \left((F'(S_j^k))^2 + (y_j^k - t_j^k) F''(S_j^k) \right) (a_j^k)^2}.$$

Учитывая, что

$$\frac{\partial^2 E_S}{\partial \alpha^2} = \sum_k \sum_j \left((F'(S_j^k))^2 + (y_j^k - t_j^k) F''(S_j^k) \right) (a_j^k)^2 > 0$$

, то при таком α , достигается минимальное значение среднеквадратичной ошибки.

Так как

$$\frac{\partial E_S}{\partial \omega_{ij}(t)} = \frac{1}{2} \frac{\partial}{\partial \omega_{ij}(t)} \sum_p \sum_q (y_q^p - t_q^p)^2 =$$

$$= \frac{1}{2} \sum_p \frac{\partial}{\partial \omega_{ij}(t)} (y_j^p - t_j^p)^2 = \sum_p (y_j^p - t_j^p) \frac{\partial}{\partial \omega_{ij}(t)} (y_j^p - t_j^p) =$$

$$= \sum_p (y_j^p - t_j^p) \frac{\partial}{\partial \omega_{ij}(t)} \left(F \left(\sum_i \omega_{ij} x_i^p - T_j \right) - t_j^p \right) =$$

$$= \sum_p (y_j^p - t_j^p) \cdot F'(S_j^p) x_i^p, \quad (i = \overline{1, n}, \quad j = \overline{1, m}) \quad (8)$$

и

$$\frac{\partial E_S}{\partial T_j(t)} = \frac{1}{2} \frac{\partial}{\partial T_j(t)} \sum_p \sum_q (y_q^p - t_q^p)^2 =$$

$$= \frac{1}{2} \sum_p \frac{\partial}{\partial T_j(t)} (y_j^p - t_j^p)^2 = \sum_p (y_j^p - t_j^p) \frac{\partial}{\partial T_j(t)} (y_j^p - t_j^p) =$$

$$= \sum_p (y_j^p - t_j^p) \frac{\partial}{\partial T_j(t)} \left(F \left(\sum_i \omega_{ij} x_i^p - T_j \right) - t_j^p \right) =$$

$$= - \sum_p (y_j^p - t_j^p) F'(S_j^p), \quad (j = \overline{1, m}) \quad (9)$$

то, подставляя эти соотношения в (5), имеем:

$$a_j^k = \sum_i \frac{\partial E_S}{\partial \omega_{ij}(t)} x_i^k - \frac{\partial E_S}{\partial T_j(t)} =$$

$$= \sum_i \sum_p (y_j^p - t_j^p) F'(S_j^p) x_i^p x_i^k + \sum_p (y_j^p - t_j^p) F'(S_j^p) =$$

$$= \sum_p (y_j^p - t_j^p) F'(S_j^p) \sum_i x_i^p x_i^k + \sum_p F'(S_j^p) (y_j^p - t_j^p) =$$

$$= \sum_p (y_j^p - t_j^p) F'(S_j^p) \left(\sum_i x_i^p x_i^k + 1 \right). \quad (10)$$

Таким образом

$$\alpha(t) = \frac{\sum_{k=1}^L \sum_{j=1}^m (y_j^k(t) - t_j^k) F'(S_j^k(t)) a_j^k(t)}{\sum_{k=1}^L \sum_{j=1}^m \left((F'(S_j^k(t)))^2 + (y_j^k(t) - t_j^k) F''(S_j^k(t)) (a_j^k(t))^2 \right)}$$

$$\text{где } a_j^k(t) = \sum_{p=1}^L (y_j^p(t) - t_j^p) F'(S_j^p(t)) \left(\sum_{i=1}^n x_i^p x_i^k + 1 \right),$$

$$j = \overline{1, m}, \quad k = \overline{1, L}.$$

Учитывая (8), (9), получим выражения для модификации синаптических связей с использованием адаптивного шага обучения (2) и (3) соответственно:

$$\omega_{ij}(t+1) = \omega_{ij}(t) - \alpha(t) \frac{\partial E_S}{\partial \omega_{ij}(t)} =$$

$$= \omega_{ij}(t) - \alpha(t) \cdot \sum_{k=1}^L (y_j^k(t) - t_j^k) \cdot F'(S_j^k(t)) x_i^k,$$

$$i = \overline{1, n}, \quad j = \overline{1, m}$$

$$T_j(t+1) = T_j(t) - \alpha(t) \frac{\partial E_S}{\partial T_j(t)} =$$

$$= T_j(t) + \alpha(t) \cdot \sum_{k=1}^L (y_j^k(t) - t_j^k) F'(S_j^k(t)), \quad j = \overline{1, m}.$$

Теорема доказана.

3. ВЫРАЖЕНИЕ ОПТИМАЛЬНОЙ ВЕЛИЧИНЫ ШАГА ОБУЧЕНИЯ С ИСПОЛЬЗОВАНИЕМ МЕТОДА НАИСКОРЕЙШЕГО СПУСКА

Рассмотрим функцию $E_S(t)$ как функцию нескольких переменных:

$$E_S(\omega_{11}, \omega_{21}, \dots, \omega_{n1}, T_1, \omega_{12}, \omega_{22}, \dots, \omega_{n2}, T_2, \dots, \omega_{1m}, \omega_{2m}, \dots, \omega_{nm}, T_m).$$

Введем обозначения:

вектора переменных

$$\bar{W} = (\omega_{11}, \omega_{21}, \dots, \omega_{n1}, T_1, \omega_{12}, \omega_{22}, \dots, \omega_{n2}, T_2, \dots, \omega_{1m}, \omega_{2m}, \dots, \omega_{nm}, T_m)^T,$$

вектора градиента функции E_S

$$\nabla E_S = \left(\frac{\partial E_S}{\partial \omega_{11}}, \frac{\partial E_S}{\partial \omega_{21}}, \dots, \frac{\partial E_S}{\partial \omega_{n1}}, \frac{\partial E_S}{\partial T_1}, \frac{\partial E_S}{\partial \omega_{12}}, \frac{\partial E_S}{\partial \omega_{22}}, \dots, \frac{\partial E_S}{\partial \omega_{n2}}, \frac{\partial E_S}{\partial T_2}, \dots, \frac{\partial E_S}{\partial \omega_{1m}}, \frac{\partial E_S}{\partial \omega_{2m}}, \dots, \frac{\partial E_S}{\partial \omega_{nm}}, \frac{\partial E_S}{\partial T_m} \right)^T$$

и матрицы Гессе вторых производных $\nabla^2 E_S$.

Разложим функцию в ряд Тейлора, ограничиваясь частными производными второго порядка включительно:

$$E_S(t+1) = E_S(t) + (\nabla E_S(t), \bar{W}(t+1) - \bar{W}(t)) +$$

$$+ \frac{1}{2} (\nabla^2 E_S(t) \cdot (\bar{W}(t+1) - \bar{W}(t)), \bar{W}(t+1) - \bar{W}(t)),$$

где $(\nabla E_S(t), \bar{W}(t+1) - \bar{W}(t))$ – скалярное произведениевекторов $\nabla E_S(t)$ и $\bar{W}(t+1) - \bar{W}(t)$,
 $(\nabla^2 E_S(t) \cdot (\bar{W}(t+1) - \bar{W}(t)), \bar{W}(t+1) - \bar{W}(t))$ – скалярное произведениевекторов $\nabla^2 E_S(t) \cdot (\bar{W}(t+1) - \bar{W}(t))$ и $(\bar{W}(t+1) - \bar{W}(t))$, а
 $\nabla^2 E_S(t) \cdot (\bar{W}(t+1) - \bar{W}(t))$ – произведение матрицы
 $\nabla^2 E_S(t)$ на вектор $(\bar{W}(t+1) - \bar{W}(t))$.

Учитывая, что в соответствии с идеей метода наискорейшего спуска

$$\bar{W}(t+1) = \bar{W}(t) - \alpha(t) \cdot \nabla E_S(t),$$

получим

$$E_S(t+1) = E_S(t) + (\nabla E_S(t), -\alpha(t) \cdot \nabla E_S(t)) +$$

$$+ \frac{1}{2} (\nabla^2 E_S(t) \cdot (-\alpha(t) \cdot \nabla E_S(t)), -\alpha(t) \cdot \nabla E_S(t))$$

или

$$E_S(t+1) = E_S(t) - \alpha(t) \cdot (\nabla E_S(t), \nabla E_S(t)) +$$

$$+ \frac{1}{2} \alpha^2(t) \cdot (\nabla^2 E_S(t) \cdot \nabla E_S(t), \nabla E_S(t)).$$

Тогда

$$\frac{\partial E_S}{\partial \alpha} = -(\nabla E_S(t), \nabla E_S(t)) +$$

$$+ \alpha(t) \cdot (\nabla^2 E_S(t) \cdot \nabla E_S(t), \nabla E_S(t)) = 0,$$

если

$$\alpha(t) = \frac{(\nabla E_S(t), \nabla E_S(t))}{(\nabla^2 E_S(t) \cdot \nabla E_S(t), \nabla E_S(t))}$$

или

$$\alpha(t) = \frac{\|\nabla E_S(t)\|^2}{(\nabla^2 E_S(t) \cdot \nabla E_S(t), \nabla E_S(t))}. \quad (11)$$

Таким образом соотношения (1), (8), (9) можно получить, используя выражение оптимальной величины шага для метода наискорейшего спуска (12).

Используя последнее соотношение, получим (1).

Действительно, учитывая что

$$\frac{\partial E_S}{\partial \omega_{ij}} = \sum_k (y_j^k - t_j^k) F'(S_j^k) x_i^k \text{ и}$$

$$\frac{\partial E_S}{\partial T_j} = -\sum_k (y_j^k - t_j^k) F'(S_j^k), \text{ имеем:}$$

$$\begin{aligned} \frac{\partial^2 E_S}{\partial \omega_{ij} \partial \omega_{qp}} &= \frac{\partial}{\partial \omega_{qp}} \sum_k (y_j^k - t_j^k) F'(S_j^k) x_i^k = \\ &= \frac{\partial}{\partial \omega_{qp}} \sum_k \left(F \left(\sum_i \omega_{ij} x_i^k - T_j \right) - t_j^k \right) F'(S_j^k) x_i^k = \\ &= \sum_k \left(F'(S_j^k) \right)^2 x_i^k x_q^k + \sum_k (y_j^k - t_j^k) F''(S_j^k) x_i^k x_q^k = \\ &= \sum_k \left(\left(F'(S_j^k) \right)^2 + (y_j^k - t_j^k) F''(S_j^k) \right) x_i^k x_q^k, \end{aligned}$$

если $j = p$, и $\frac{\partial^2 E_S}{\partial \omega_{ij} \partial \omega_{qp}} = 0$ в противном случае;

$$\begin{aligned} \frac{\partial^2 E_S}{\partial \omega_{ij} \partial T_m} &= \frac{\partial}{\partial T_m} \sum_k (y_j^k - t_j^k) F'(S_j^k) x_i^k = \\ &= \frac{\partial}{\partial T_m} \left(\sum_k \left(F \left(\sum_i \omega_{ij} x_i^k - T_j \right) - t_j^k \right) F'(S_j^k) x_i^k \right) = \\ &= -\sum_k \left(\left(F'(S_j^k) \right)^2 + (y_j^k - t_j^k) F''(S_j^k) \right) x_i^k, \end{aligned}$$

если $j = m$, и $\frac{\partial^2 E_S}{\partial \omega_{ij} \partial T_m} = 0$ в противном случае;

$$\begin{aligned} \frac{\partial^2 E_S}{\partial T_j \partial T_m} &= \frac{\partial}{\partial T_m} \left(-\sum_k (y_j^k - t_j^k) F'(S_j^k) \right) = \\ &= \frac{\partial}{\partial T_m} \left(-\sum_k \left(F \left(\sum_i \omega_{ij} x_i^k - T_j \right) - t_j^k \right) F'(S_j^k) \right) = \\ &= \sum_k \left(\left(F'(S_j^k) \right)^2 + (y_j^k - t_j^k) F''(S_j^k) \right), \end{aligned}$$

если $j = m$, и $\frac{\partial^2 E_S}{\partial T_j \partial T_m} = 0$ в противном случае.

Тогда получим (12')

Вычислим отдельно слагаемые числителя:

$$\begin{aligned} \sum_i \sum_j \left(\frac{\partial E_S}{\partial \omega_{ij}} \right)^2 &= \sum_i \sum_j \left(\sum_k (y_j^k - t_j^k) F'(S_j^k) x_i^k \right)^2 = \\ &= \sum_i \sum_j \left(\sum_k (y_j^k - t_j^k) F'(S_j^k) x_i^k \right) \left(\sum_p (y_j^p - t_j^p) F'(S_j^p) x_i^p \right) = \\ &= \sum_i \sum_j \left(\sum_k (y_j^k - t_j^k) F'(S_j^k) \sum_p (y_j^p - t_j^p) F'(S_j^p) x_i^k x_i^p \right) = \\ &= \sum_j \sum_k (y_j^k - t_j^k) F'(S_j^k) \left(\sum_p (y_j^p - t_j^p) F'(S_j^p) \sum_i x_i^k x_i^p \right) \\ \text{и} \\ \sum_j \left(\frac{\partial E_S}{\partial T_j} \right)^2 &= \sum_j \left(-\sum_k (y_j^k - t_j^k) F'(S_j^k) \right)^2 = \\ &= \sum_j \left(\sum_k (y_j^k - t_j^k) F'(S_j^k) \right) \left(\sum_p (y_j^p - t_j^p) F'(S_j^p) \right) = \end{aligned}$$

$$\alpha(t) = \frac{\sum_i \sum_j \left(\frac{\partial E_S}{\partial \omega_{ij}} \right)^2 + \sum_j \left(\frac{\partial E_S}{\partial T_j} \right)^2}{\sum_i \sum_j \sum_q \sum_p \frac{\partial^2 E_S}{\partial \omega_{ij} \partial \omega_{qp}} \frac{\partial E_S}{\partial \omega_{ij}} \frac{\partial E_S}{\partial \omega_{qp}} + 2 \cdot \sum_i \sum_j \sum_m \frac{\partial^2 E_S}{\partial \omega_{ij} \partial T_m} \frac{\partial E_S}{\partial \omega_{ij}} \frac{\partial E_S}{\partial T_m} + \sum_j \sum_m \frac{\partial^2 E_S}{\partial T_j \partial T_m} \frac{\partial E_S}{\partial T_j} \frac{\partial E_S}{\partial T_m}}. \quad (12)$$

$$\alpha(t) = \frac{\sum_i \sum_j \left(\frac{\partial E_S}{\partial \omega_{ij}} \right)^2 + \sum_j \left(\frac{\partial E_S}{\partial T_j} \right)^2}{\sum_j \sum_i \sum_q \frac{\partial^2 E_S}{\partial \omega_{ij} \partial \omega_{jq}} \frac{\partial E_S}{\partial \omega_{ij}} \frac{\partial E_S}{\partial \omega_{jq}} + 2 \sum_j \sum_i \frac{\partial^2 E_S}{\partial \omega_{ij} \partial T_j} \frac{\partial E_S}{\partial \omega_{ij}} \frac{\partial E_S}{\partial T_j} + \sum_j \frac{\partial^2 E_S}{\partial T_j^2} \left(\frac{\partial E_S}{\partial T_j} \right)^2} \quad (12')$$

$$\alpha(t) = \frac{\sum_{j=1}^m \left(F'(S_j) \right)^2 (y_j - t_j)^2}{\left(\sum_{i=1}^n x_i^2 + 1 \right) \sum_{j=1}^m \left(\left(F'(S_j) \right)^2 + (y_j - t_j) F''(S_j) \right) \left(F'(S_j) \right)^2 (y_j - t_j)^2} \quad (14)$$

$$= \sum_j \sum_k (y_j^k - t_j^k) F'(S_j^k) \left(\sum_p (y_j^p - t_j^p) F'(S_j^p) \right).$$

Тогда числитель равен

$$\begin{aligned} \sum_i \sum_j \left(\frac{\partial E_S}{\partial \omega_{ij}} \right)^2 + \sum_j \left(\frac{\partial E_S}{\partial T_j} \right)^2 &= \\ &= \sum_j \sum_k (y_j^k - t_j^k) F'(S_j^k) \left(\sum_p (y_j^p - t_j^p) F'(S_j^p) \left(\sum_i x_i^p x_i^k + 1 \right) \right) = \\ &= \sum_j \sum_k (y_j^k - t_j^k) F'(S_j^k) h_k^j, \end{aligned}$$

учитывая (10).

Вычисляя слагаемые знаменателя, получим

$$\begin{aligned} \sum_j \sum_i \sum_q \frac{\partial^2 E_S}{\partial \omega_{ij} \partial \omega_{qj}} \frac{\partial E_S}{\partial \omega_{ij}} \frac{\partial E_S}{\partial \omega_{qj}} &= \\ &= \sum_i \sum_j \sum_q \left(\sum_k \left((F'(S_j^k))^2 + (y_j^k - t_j^k) F''(S_j^k) \right) x_i^k x_q^k \right) \times \\ &\times \left(\sum_p (F'(S_j^p)) (y_j^p - t_j^p) x_i^p \right) \left(\sum_m (F'(S_j^m)) (y_j^m - t_j^m) x_q^m \right) = \\ &= \sum_i \sum_j \sum_q \sum_k \left((F'(S_j^k))^2 + (y_j^k - t_j^k) F''(S_j^k) \right) \times \\ &\times \left(\sum_p (F'(S_j^p)) (y_j^p - t_j^p) x_i^k x_i^p \right) \left(\sum_m (F'(S_j^m)) (y_j^m - t_j^m) x_q^k x_q^m \right) = \\ &= \sum_j \sum_k \left((F'(S_j^k))^2 + (y_j^k - t_j^k) F''(S_j^k) \right) \times \\ &\times \left(\sum_p (F'(S_j^p)) (y_j^p - t_j^p) \sum_i x_i^k x_i^p \right) \times \\ &\times \left(\sum_m (y_j^m - t_j^m) (F'(S_j^m)) \sum_q x_q^k x_q^m \right) = \\ &= \sum_j \sum_k \left((F'(S_j^k))^2 + (y_j^k - t_j^k) F''(S_j^k) \right) \times \\ &\times \left(\sum_p (F'(S_j^p)) (y_j^p - t_j^p) \sum_i x_i^k x_i^p \right)^2, \\ 2 \cdot \sum_i \sum_j \frac{\partial^2 E_S}{\partial \omega_{ij} \partial T_j} \frac{\partial E_S}{\partial \omega_{ij}} \frac{\partial E_S}{\partial T_j} &= \\ &= 2 \sum_i \sum_j \left(- \sum_k \left((F'(S_j^k))^2 + (y_j^k - t_j^k) F''(S_j^k) \right) x_i^k \right) \times \\ &\times \left(\sum_p (F'(S_j^p)) (y_j^p - t_j^p) x_i^p \right) \left(- \sum_m (F'(S_j^m)) (y_j^m - t_j^m) \right) = \end{aligned}$$

$$\begin{aligned} &= 2 \sum_j \sum_k \left((F'(S_j^k))^2 + (y_j^k - t_j^k) F''(S_j^k) \right) \times \\ &\times \left(\sum_p (F'(S_j^p)) (y_j^p - t_j^p) \sum_i x_i^k x_i^p \right) \left(\sum_m (F'(S_j^m)) (y_j^m - t_j^m) \right), \\ \sum_j \frac{\partial^2 E_S}{\partial T_j^2} \left(\frac{\partial E_S}{\partial T_j} \right)^2 &= \\ &= \sum_j \sum_k \left((F'(S_j^k))^2 + (y_j^k - t_j^k) F''(S_j^k) \right) \times \\ &\times \left(\sum_m (F'(S_j^m)) (y_j^m - t_j^m) \right)^2 \end{aligned}$$

Тогда знаменатель равен

$$\begin{aligned} \sum_j \sum_i \sum_q \frac{\partial^2 E_S}{\partial \omega_{ij} \partial \omega_{qj}} \frac{\partial E_S}{\partial \omega_{ij}} \frac{\partial E_S}{\partial \omega_{qj}} &+ \\ + 2 \cdot \sum_i \sum_j \frac{\partial^2 E_S}{\partial \omega_{ij} \partial T_j} \frac{\partial E_S}{\partial \omega_{ij}} \frac{\partial E_S}{\partial T_j} &+ \sum_j \frac{\partial^2 E_S}{\partial T_j^2} \left(\frac{\partial E_S}{\partial T_j} \right)^2 = \\ &= \sum_j \sum_k \left((F'(S_j^k))^2 + (y_j^k - t_j^k) F''(S_j^k) \right) \times \\ &\times \left(\sum_p (F'(S_j^p)) (y_j^p - t_j^p) \sum_i x_i^k x_i^p \right)^2 \\ &+ 2 \sum_j \sum_k \left((F'(S_j^k))^2 + (y_j^k - t_j^k) F''(S_j^k) \right) \times \\ &\times \left(\sum_p (F'(S_j^p)) (y_j^p - t_j^p) \sum_i x_i^k x_i^p \right) \left(\sum_m (F'(S_j^m)) (y_j^m - t_j^m) \right) + \\ &+ \sum_j \sum_k \left((F'(S_j^k))^2 + (y_j^k - t_j^k) F''(S_j^k) \right) \times \\ &\times \left(\sum_m (F'(S_j^m)) (y_j^m - t_j^m) \right)^2 = \\ &= 2 \left(\left(\sum_p (F'(S_j^p)) (y_j^p - t_j^p) \sum_i x_i^k x_i^p \right) \times \right. \\ &\times \left(\sum_m (F'(S_j^m)) (y_j^m - t_j^m) \right) + \left. \left(\sum_m (F'(S_j^m)) (y_j^m - t_j^m) \right)^2 \right) = \\ &= \sum_j \sum_k \left((F'(S_j^k))^2 + (y_j^k - t_j^k) F''(S_j^k) \right) \times \\ &\times \left(\sum_p (F'(S_j^p)) (y_j^p - t_j^p) \sum_i x_i^k x_i^p + \sum_m (F'(S_j^m)) (y_j^m - t_j^m) \right)^2 = \\ &= \sum_j \sum_k \left((F'(S_j^k))^2 + (y_j^k - t_j^k) F''(S_j^k) \right) \times \end{aligned}$$

$$\times \left(\sum_p (F'(S_j^p)) (y_j^p - t_j^p) \left(\sum_i x_i^k x_i^p + 1 \right) \right)^2 = \\ = \sum_j \sum_k \left((F'(S_j^k))^2 + (y_j^k - t_j^k) F''(S_j^k) \right) (a_j^k)^2,$$

учитывая (10).

Таким образом

$$\alpha(t) = \frac{\sum_{k=1}^L \sum_{j=1}^m F'(S_j^k) (y_j^k - t_j^k) a_j^k}{\sum_{k=1}^L \sum_{j=1}^m \left((F'(S_j^k))^2 + (y_j^k - t_j^k) F''(S_j^k) \right) (a_j^k)^2},$$

$$\text{где } a_j^k = \sum_{p=1}^L (y_j^p - t_j^p) F'(S_j^p) \left(\sum_{i=1}^n x_i^p x_i^k + 1 \right),$$

$$j = \overline{1, m}, \quad k = \overline{1, L}.$$

Получим частные случаи соотношения (1), сформулированные в виде следствий.

Следствие 1. В случае одного j -ого выходного нейронного элемента, соотношение (1), примет вид:

$$\alpha(t) = \frac{\sum_{k=1}^L F'(S_j^k) (y_j^k - t_j^k) a_j^k}{\sum_{k=1}^L \left((F'(S_j^k))^2 + (y_j^k - t_j^k) F''(S_j^k) \right) (a_j^k)^2}, \quad (13)$$

$$\text{где } a_j^k = \sum_{p=1}^L (y_j^p - t_j^p) F'(S_j^p) \left(\sum_{i=1}^n x_i^p x_i^k + 1 \right), \quad k = \overline{1, L}.$$

Следствие 2. В случае одного образа, т.е. при $L = 1$, соотношение (1), принимает вид:

$$\alpha(t) = \frac{\sum_{j=1}^m F'(S_j) (y_j - t_j) a_j}{\sum_{j=1}^m \left((F'(S_j))^2 + (y_j - t_j) F''(S_j) \right) (a_j)^2},$$

$$\text{где } a_j = (y_j - t_j) F'(S_j) \left(\sum_{i=1}^n x_i^2 + 1 \right), \quad j = \overline{1, m}, \quad \text{или}$$

определяется соотношением (14).

Заметим, что соотношение (12) может быть использовано для получения выражений величины адаптивного шага обучения и для других функций E_S , например,

$$E_S = \sum_k \sum_j |y_j^k - t_j^k|.$$

4. ОЦЕНКИ ШАГА ОБУЧЕНИЯ НЕЙРОННОЙ СЕТИ

Получим оценку величины адаптивного шага обучения.

Рассмотрим матрицу Гессе функции E_S . Учитывая выражения для частных производных второго порядка, полученные выше, имеем

$$\nabla^2 E_S = \begin{pmatrix} \nabla^2 E_1 & 0 & \dots & 0 \\ 0 & \nabla^2 E_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \nabla^2 E_m \end{pmatrix}_{m(n+1) \times m(n+1)} \quad (15)$$

где

$$\nabla^2 E_j = \begin{pmatrix} \frac{\partial^2 E}{\partial \omega_{1j}^2} & \frac{\partial^2 E}{\partial \omega_{1j} \partial \omega_{2j}} & \dots & \frac{\partial^2 E}{\partial \omega_{1j} \partial \omega_{nj}} & \frac{\partial^2 E}{\partial \omega_{1j} \partial T_j} \\ \frac{\partial^2 E}{\partial \omega_{2j} \partial \omega_{1j}} & \frac{\partial^2 E}{\partial \omega_{2j}^2} & \dots & \frac{\partial^2 E}{\partial \omega_{2j} \partial \omega_{nj}} & \frac{\partial^2 E}{\partial \omega_{2j} \partial T_j} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \frac{\partial^2 E}{\partial \omega_{nj} \partial \omega_{1j}} & \frac{\partial^2 E}{\partial \omega_{nj} \partial \omega_{2j}} & \dots & \frac{\partial^2 E}{\partial \omega_{nj}^2} & \frac{\partial^2 E}{\partial \omega_{nj} \partial T_j} \\ \frac{\partial^2 E}{\partial T_j \partial \omega_{1j}} & \frac{\partial^2 E}{\partial T_j \partial \omega_{2j}} & \dots & \frac{\partial^2 E}{\partial T_j \partial \omega_{nj}} & \frac{\partial^2 E}{\partial T_j^2} \end{pmatrix}_{(n+1) \times (n+1)}$$

Тогда в случае одного образа

$$\nabla^2 E_j(t) = \left((F'(S_j))^2 + (y_j - t_j) F''(S_j) \right) \times$$

$$\times \begin{pmatrix} x_1^2 & x_1 x_2 & \dots & x_1 x_n & -x_1 \\ x_2 x_1 & x_2^2 & \dots & x_2 x_n & -x_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ x_n x_1 & x_n x_2 & \dots & x_n^2 & -x_n \\ -x_1 & -x_2 & \dots & -x_n & 1 \end{pmatrix}.$$

Заметим, что строки и столбцы матрицы $\nabla^2 E_j(t)$ пропорциональны. Поэтому матрица $\nabla^2 E_j(t)$ является неотрицательно определенной, так как главные ее миноры Δ_i ($i = \overline{1, n+1}$) неотрицательны

$$\Delta_1 = \left((F'(S_j))^2 + (y_j - t_j) F''(S_j) \right) x_1^2 \geq 0,$$

$$\Delta_i = 0, \quad (i = \overline{2, n+1}).$$

Найдем собственные значения матрицы $\nabla^2 E_j(t)$:

$$\begin{aligned} |\nabla^2 E_j(t) - \lambda I| &= \\ &= \begin{vmatrix} cx_1^2 - \lambda & cx_1 x_2 & \dots & cx_1 x_n & -cx_1 \\ cx_2 x_1 & cx_2^2 - \lambda & \dots & cx_2 x_n & -cx_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ cx_n x_1 & cx_n x_2 & \dots & cx_n^2 - \lambda & -cx_n \\ -cx_1 & -cx_2 & \dots & -cx_n & c - \lambda \end{vmatrix} = \\ &= \begin{vmatrix} cx_1^2 - \lambda & cx_1 x_2 & \dots & cx_1 x_n & -cx_1 \\ \frac{x_2}{x_1} \lambda & -\lambda & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \frac{x_n}{x_1} \lambda & 0 & \dots & -\lambda & 0 \\ -\frac{1}{x_1} \lambda & 0 & \dots & 0 & -\lambda \end{vmatrix} = \end{aligned}$$

$$\begin{aligned}
&= (-1)^n \lambda^n \begin{vmatrix} cx_1^2 - \lambda & cx_1x_2 & \cdots & cx_1x_n & -cx_1 \\ -\frac{x_2}{x_1} & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ -\frac{x_n}{x_1} & 0 & \cdots & 1 & 0 \\ \frac{1}{x_1} & 0 & \cdots & 0 & 1 \end{vmatrix} = \\
&= (-1)^n \lambda^n \begin{vmatrix} c\left(\sum_i x_i^2 + 1\right) - \lambda & 0 & \cdots & 0 & 0 \\ -\frac{x_2}{x_1} & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ -\frac{x_n}{x_1} & 0 & \cdots & 1 & 0 \\ \frac{1}{x_1} & 0 & \cdots & 0 & 1 \end{vmatrix} = \\
&= (-1)^n \lambda^n \left(c\left(\sum_i x_i^2 + 1\right) - \lambda \right),
\end{aligned}$$

где

$$c = (F'(S_j))^2 + (y_j - t_j)F''(S_j).$$

Тогда

$$\begin{aligned}
\lambda_j &= c \left(\sum_i x_i^2 + 1 \right) = \\
&= \left((F'(S_j))^2 + (y_j - t_j)F''(S_j) \right) \left(\sum_i x_i^2 + 1 \right)
\end{aligned}$$

является наибольшим собственным значением $\nabla^2 E_j(t)$.

Поэтому, норма матрицы $\nabla^2 E_j(t)$, связанная с метрикой $\|x\| = \sqrt{(x, x)}$ определяется соотношением

$$\|\nabla^2 E_j(t)\| = \left((F'(S_j))^2 + (y_j - t_j)F''(S_j) \right) \left(\sum_i x_i^2 + 1 \right).$$

Следовательно, для такой же нормы матрицы $\nabla^2 E_S(t)$, учитывая (15), выполняется

$$\begin{aligned}
\|\nabla^2 E_S(t)\| &= \max_j \|\nabla^2 E_j(t)\| = \\
&= \left(\sum_i x_i^2 + 1 \right) \cdot \max_j \left((F'(S_j))^2 + (y_j - t_j)F''(S_j) \right).
\end{aligned}$$

Учитывая, что для скалярного произведения верно неравенство

$$(\nabla^2 E_S(t) \cdot \nabla E_S(t), \nabla E_S(t)) \leq \|\nabla^2 E_S(t)\| \cdot \|\nabla E_S(t)\|^2,$$

имеем

$$\begin{aligned}
\alpha(t) &= \frac{\|\nabla E_S(t)\|^2}{(\nabla^2 E_S(t) \cdot \nabla E_S(t), \nabla E_S(t))} \geq \\
&\geq \frac{\|\nabla E_S(t)\|^2}{\|\nabla^2 E_S(t)\| \cdot \|\nabla E_S(t)\|^2} = \frac{1}{\|\nabla^2 E_S(t)\|}.
\end{aligned}$$

Таким образом, имеет место следующая оценка адаптивного шага обучения:

$$\alpha(t) \geq \frac{1}{\left(\sum_i x_i^2 + 1 \right) \cdot \max_j \left((F'(S_j))^2 + (y_j - t_j)F''(S_j) \right)}. \quad (16)$$

Рассмотрим случай группового обучения. Введем обозначения

$$\begin{aligned}
\nabla^2 E_j^k(t) &= \left((F'(S_j^k))^2 + (y_j^k - t_j^k)F''(S_j^k) \right) \times \\
&\times \begin{pmatrix} x_1^{k^2} & x_1^k x_2^k & \cdots & x_1^k x_n^k & -x_1^k \\ x_2^k x_1^k & x_2^{k^2} & \cdots & x_2^k x_n^k & -x_2^k \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ x_n^k x_1^k & x_n^k x_2^k & \cdots & x_n^{k^2} & -x_n^k \\ -x_1^k & -x_2^k & \cdots & -x_n^k & 1 \end{pmatrix}.
\end{aligned}$$

Тогда, учитывая, что

$$\begin{aligned}
\frac{\partial^2 E_S}{\partial \omega_{ij} \partial \omega_{qj}} &= \sum_k \left((F'(S_j^k))^2 + (y_j^k - t_j^k)F''(S_j^k) \right) x_i^k x_q^k, \\
\frac{\partial^2 E_S}{\partial \omega_{ij} \partial T_j} &= - \sum_k \left((F'(S_j^k))^2 + (y_j^k - t_j^k)F''(S_j^k) \right) x_i^k, \\
\frac{\partial^2 E_S}{\partial T_j^2} &= \sum_k \left((F'(S_j^k))^2 + (y_j^k - t_j^k)F''(S_j^k) \right),
\end{aligned}$$

имеем:

$$\nabla^2 E_j(t) = \sum_k \nabla^2 E_j^k(t).$$

Так как для любого k выполняется $\nabla^2 E_j^k(t) \geq 0$, то

$$\max_k \|\nabla^2 E_j^k(t)\| \leq \|\nabla^2 E_j(t)\| \leq \sum_k \|\nabla^2 E_j^k(t)\|.$$

Следовательно,

$$\begin{aligned}
\max_k \left((F'(S_j^k))^2 + (y_j^k - t_j^k)F''(S_j^k) \right) \left(\sum_i x_i^{k^2} + 1 \right) &\leq \\
\leq \|\nabla^2 E_j(t)\| &\leq \sum_k \left((F'(S_j^k))^2 + (y_j^k - t_j^k)F''(S_j^k) \right) \left(\sum_i x_i^{k^2} + 1 \right).
\end{aligned}$$

Тогда, учитывая (15), имеем

$$\begin{aligned}
\|\nabla^2 E_S(t)\| &= \max_j \|\nabla^2 E_j(t)\| \leq \\
&\leq \max_j \sum_k \left((F'(S_j^k))^2 + (y_j^k - t_j^k)F''(S_j^k) \right) \left(\sum_i x_i^{k^2} + 1 \right)
\end{aligned}$$

и

$$\|\nabla^2 E_S(t)\| = \max_j \|\nabla^2 E_j(t)\| \geq$$

$$\geq \max_j \max_k \left((F'(S_j^k))^2 + (y_j^k - t_j^k) F''(S_j^k) \right) \left(\sum_i x_i^{k^2} + 1 \right)$$

Учитывая, что

$$\alpha(t) = \frac{\|\nabla E_S(t)\|^2}{(\nabla^2 E_S(t) \cdot \nabla E_S(t), \nabla E_S(t))} \geq$$

$$\geq \frac{\|\nabla E_S(t)\|^2}{\|\nabla^2 E_S(t)\| \cdot \|\nabla E_S(t)\|^2} = \frac{1}{\|\nabla^2 E_S(t)\|},$$

имеем

$$\alpha(t) \geq \frac{1}{\|\nabla^2 E_S(t)\|} \geq$$

$$\geq \frac{1}{\max_j \sum_k \left((F'(S_j^k))^2 + (y_j^k - t_j^k) F''(S_j^k) \right) \left(\sum_i x_i^{k^2} + 1 \right)}.$$

Теорема 2. Для адаптивного шага обучения, определяемого соотношением (1), верна следующая оценка снизу:

$$\alpha(t) \geq \frac{1}{\max_j \sum_k \left((F'(S_j^k))^2 + (y_j^k - t_j^k) F''(S_j^k) \right) \left(\sum_i x_i^{k^2} + 1 \right)}. \quad (17)$$

5. АЛГОРИТМ ОБУЧЕНИЯ НЕЙРОННОЙ СЕТИ И ЕГО МОДИФИКАЦИИ

Приведем алгоритм обучения нейронной сети, использующий соотношения (1)-(4):

1. Задается минимальная среднеквадратичная ошибка сети E_m , которой необходимо достичь в процессе обучения.

2. Случайным образом инициализируются весовые коэффициенты сети ω_{ij} ($i = \overline{1, n}$, $j = \overline{1, m}$), и пороговые значения нейронных элементов T_j ($j = \overline{1, m}$).

3. Подаются входные образы $\overline{x^k} = (x_1^k, \dots, x_n^k)$ ($k = \overline{1, L}$) на нейронную сеть и вычисляются векторы $\overline{y^k} = (y_1^k, \dots, y_m^k)$ ($k = \overline{1, L}$) выходной активности сети, определяемые соотношениями (4).

4. Вычисляется величина адаптивного шага обучения $\alpha(t)$, определяемая соотношением (1).

5. Производится изменение весовых коэффициентов $\omega_{ij}(t+1)$ ($i = \overline{1, n}$, $j = \overline{1, m}$) и порогов нейронной сети $T_j(t+1)$ ($j = \overline{1, m}$) согласно выражениям (2) и (3), соответственно.

6. Алгоритм завершает свою работу, если суммарная среднеквадратичная ошибка сети

$$E_S = \frac{1}{2} \sum_{j=1}^m \sum_{k=1}^L (y_j^k - t_j^k)^2$$

не превосходит заданной величины E_m , т.е. $E_S \geq E_m$, в противном случае выполняется п. 2.

В зависимости от решаемой задачи (п. 4, 5 алгоритма) можно комбинировать следующие методы изменения весовых коэффициентов и порогов нейронной сети:

1. *Метод координатной релаксации.* На каждом шаге “ликвидируют” наибольшее по абсолютной величине значение $E_S(j) = \frac{1}{2} \sum_{k=1}^L (y_j^k - t_j^k)^2$ среднеквадратичной ошибки

j – ого нейронного элемента выходного слоя путем исправления весовых коэффициентов $\omega_{ij}(t+1)$ ($i = \overline{1, n}$) и порога нейронной сети $T_j(t+1)$ согласно выражениям (2) и (3), с учетом соотношения (13).

2. *Метод групповой релаксации.* На каждом шаге “ликвидируют” наибольшее по абсолютной величине значение

$$E_S(k) = \frac{1}{2} \sum_{j=1}^m (y_j^k - t_j^k)^2$$

путем исправления весовых коэффициентов $\omega_{ij}(t+1)$ ($i = \overline{1, n}$, $j = \overline{1, m}$) и порогов нейронной сети $T_j(t+1)$ ($j = \overline{1, m}$) согласно выражениям (2), (3) и (14) для соответствующего k – ого образа.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Golovko V., Savitsky Ju., Gladyschuk V. Predicting Neural Net // Proceedings Int. Conf. CM NDT-D5. – Berlin: DGZfP. – 1995. – P. 348-353.
2. Головки В. А. Нейроинтеллект: теория и применение. Книга 1: Организация и обучение нейронных сетей с прямыми и обратными связями. – Брест: Изд. БПИ, 1999. – 264 с.